ELSEVIER

Assessing drug-likeness – what are we missing?

Giulio Vistoli[1], Alessandro Pedretti[1] and Bernard Testa[2]

[1] Istituto di Chimica Farmaceutica "Pietro Pratesi", Facoltà di Farmacia, Università di Milano, Via Mangiagalli 25, I-20133 Milano, Italy
[2] Department of Pharmacy, University Hospital Centre (CHUV), Rue du Bugnon, CH-1011 Lausanne, Switzerland

The concept of drug-likeness helps to optimise pharmacokinetic and pharmaceutical properties, for example, solubility, chemical stability, bioavailability and distribution profile. A number of molecular descriptors have emerged as reasonably informative and predictive, for example, the Rule-of-Five. Here, we review some current approaches, then discuss their major shortcoming, namely the static nature of the structural features and physicochemical properties they encode. As we demonstrate, molecules are not 'frozen statues' but 'dancing ballerinas', and several of their computable physicochemical properties are conformation-dependent and lead to the concept of property spaces. Molecular sensitivity (namely, how much a given computable physicochemical property varies as a function of flexibility) appears as a promising descriptor to encode some of the information contained in molecular property spaces.

Druggability and drug-likeness: a semantic introduction

The concept of druggability emerged as a way of describing members of the expressed proteome that can bind high-affinity ligands with drug-like properties, and whose activity is modulated by such ligands [1–3]. The concept was also expanded to describe the 'druggable genome', that is, those genes that code for physiologically relevant proteins whose activity can be modulated by high-affinity ligands. Because some DNA and RNA sequences may also serve as sites of action of drugs, the concept of druggability is more generally applied to targets [4]. But as the concept of druggability 'replicated itself from brain to brain' [cf. Dawkin's meme concept, Ref. 5], its meaning appears to have broadened (perhaps infelicitously) beyond its generally accepted definition to signify different things to different people along the discovery, development and clinical 'pipeline' (Fig. 1). Thus, medicinal chemists deal with hits and lead compounds, and their mission is to supply pharmacologists and clinicians with candidates that are 'drug-like'

Abbreviations: ADMET, absorption, distribution, metabolism, excretion and toxicity; log D, log of distribution coefficient (pH-dependent); log P, log of partition coefficient (for a single electrical state); MD, molecular dynamics; MEP, molecular electrostatic potential; MIF, molecular interaction field; MLP, molecular lipophilicity potential; PC, principal component; PCA, principal component analysis; PD, pharmacodynamic(s); PK, pharmacokinetic(s); PSA, polar surface area; RMSD, root mean square deviation; Ro5, Rule-of-Five; SAS, solvent-accessible surface area.

GIULIO VISTOLI

Giulio Vistoli was born in 1968. He received his Laurea degree in medicinal chemistry at the University of Milan in 1994. During his PhD studies with Prof. L. Villa, he spent a period in Lausanne under the supervision of Prof. B. Testa with whom he has fruitfully collaborated since 1996. In 1999, he became assistant professor in medicinal chemistry at University of Milan. His recent researches focus on the development of both property space concept to explore the dynamic profile of molecular fields and novel approaches to model the transmembrane proteins.

ALESSANDRO PEDRETTI

Alessandro Pedretti was born in 1970. He received his Laurea degree in medicinal chemistry at the University of Milan in 1995. After PhD studies with Prof. L. Villa, he became assistant professor in medicinal chemistry at the University of Milan in 2001. His interests mainly deal with computer programming applied in computational chemistry, realising novel software tools for molecular modelling and docking analysis. In particular, he developed VEGA (available at http://www.ddl.unimi.it), a program able to compute several molecular fields analysing their dynamic profiles during the simulation time.

BERNARD TESTA

Bernard Testa graduated in pharmacy and obtained a PhD on the physicochemistry of drug–macromolecule interactions. He spent two years as a post-doctoral fellow at the University of London, under the supervision of Prof. Arnold H. Beckett. In 1978, he became full professor and Head of Medicinal Chemistry at the University of Lausanne. He served as Dean of the Faculty of Sciences, Director of the Geneva-Lausanne School of Pharmacy, and President of the University Senate. He has written or edited 34 books (3 more are in preparation), and (co)-authored over 450 research and review articles in the fields of drug design and drug metabolism. He is a member of the Editorial Board of several major journals and of a number of international and national scientific societies. He holds Honorary Doctorates from three Universities, and was in particular a recipient of the Nauta Award on Pharmacochemistry awarded by the European Federation for Medicinal Chemistry.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.11.007

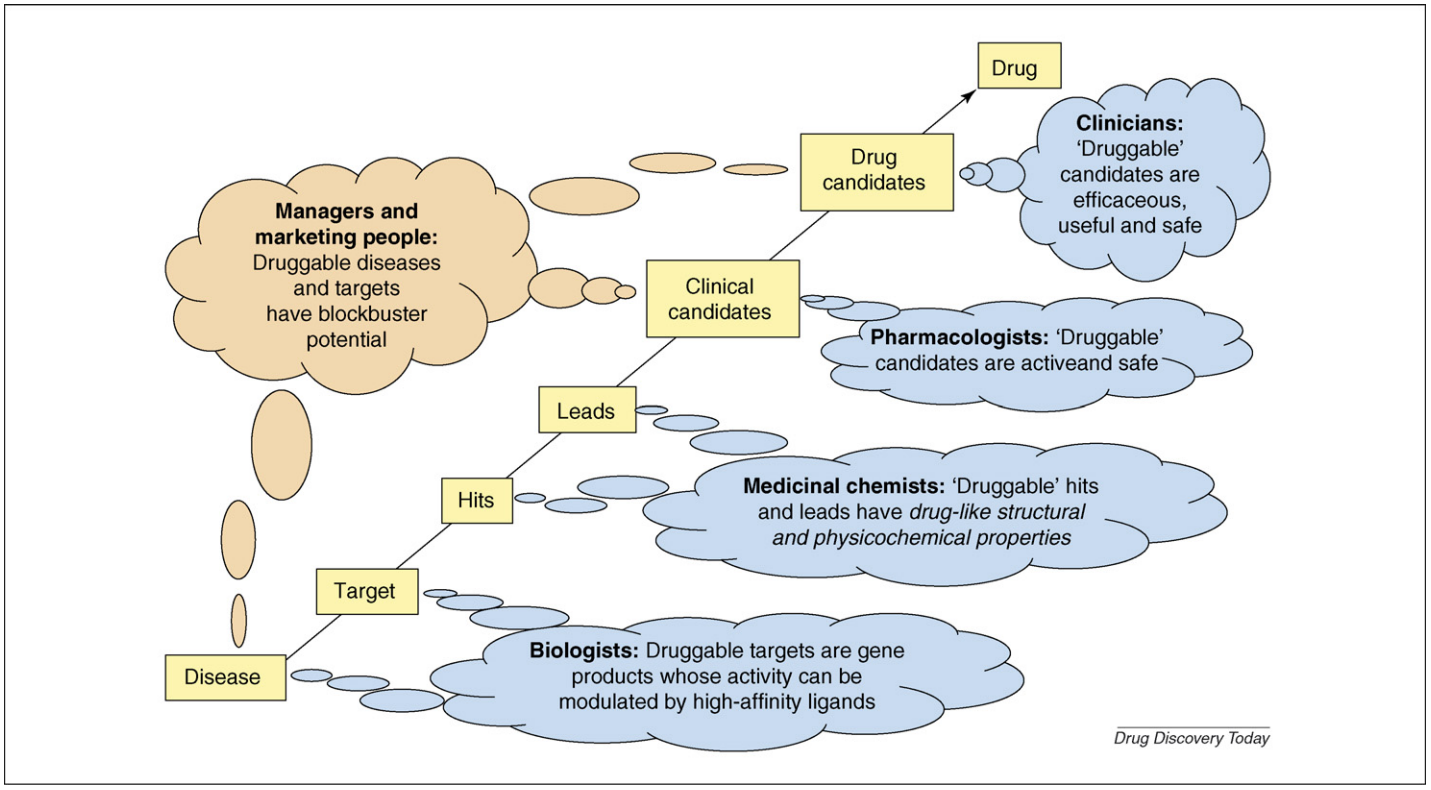

FIGURE 1

A druggable target is a protein or another functional biomacromolecule whose activity can be modulated by high-affinity ligands with drug-like properties. Perhaps infelicitously, the concept of druggability is sometimes used beyond its original meaning, as schematised here. The figure also serves to place drug-likeness, the main focus of this paper, in the broader context of drug discovery and development.

and capable of modulating targets [6]. As an aside, we note that druggability implicitly, or even explicitly, means 'blockbuster-ability' to some marketing managers, but this is of no concern here.

The mission of medicinal chemists is to design and discover hits that can be improved to leads, leads that can be optimised to candidates and candidates that will become valuable drugs. In other words, the strategy of medicinal chemists in their research is to discover new chemical entities which maximally resemble existing drugs with respect to key physicochemical and biological properties, with the knowledge that the quest for 'drug-like' properties may indeed help achieve good pharmacokinetic and pharmacodynamic properties. Such drug-like properties include lipophilicity, polarity, size and solubility, plus some biological properties such as no or low inhibition of cytochromes P450, no QT prolongation, metabolic stability, absence of toxicophoric groups.

A major challenge for medicinal chemists is, therefore, to acquire as much useful and relevant information as possible on the structural features and physicochemical properties of their compounds, be they virtual or material. As we will show, the magnitude of this challenge is largely underestimated [7]. They ignore the dynamic nature and adaptability range of molecular entities because most physicochemical profiles are generally sets of single-point descriptor values. A more specific problem faced by medicinal chemists in their search for drug-likeness occurs when the structural prerequisites for the target PD activity and desired PK behaviour are incompatible. Early identification of this obstacle can allow an '*ad hoc*' prodrug strategy to be implemented, as discussed elsewhere [8–10].

In this review, we will initially survey the descriptors applied in profiling chemical compounds, and how their use as independent variables in multivariate analyses allows chemical space to be computed. The second part of the article will demonstrate that chemical entities are not static but multidimensional objects whose 3D structure and physicochemical properties adapt to their environment. Hence single-value descriptors, despite their obvious utility, encode only part of the structural and physicochemical information of chemical compounds. This fact severely restricts their predictive capacity and justifies a far greater research effort in the structural and physicochemical characterisation of promising bioactive compounds.

The current understanding of drug-likeness

Molecular descriptors and the concept of chemical diversity space (a.k.a. chemical space)

A nice illustration of the type of descriptors used to characterise chemical compounds is afforded by the work of Feher and Schmidt [11]. They used a set of 41 descriptors and applied them to a set of 670,536 combinatorial compounds, a set of 10,968 drugs and a set of 3287 natural products. Table 1 presents a selection of 24 descriptors and their mean value in the three compound classes, showing that the values of some descriptors differ markedly between the three classes, whereas others do not. Most of the descriptors in this study were counts of various features, or ratios of counts. Only three were not able to be counted directly, but were

TABLE 1

Mean values of different structural descriptors among combichem compounds, drugs and natural compounds [11][a,b]

Parameters	Descriptors		
	Combinatorial compounds	Drugs	Natural compounds
Molecular weight	(393)	(340)	(414)
Number of chiral centres	**0.4**	**2.3**	**6.2**
Number of rotatable bonds	**(6.4)**	**(5.6)**	**(4.4)**
Ratio of aromatic atoms to ring atoms	**0.8**	**0.55**	**0.31**
Degree of unsaturation	(12.0)	(8.0)	(8.8)
Number of ring systems	(2.6)	(1.7)	(1.7)
Ring fusion degree	**1.27**	**1.67**	**2.83**
Number of C atoms	(20.5)	(17.2)	(22.8)
Number of N atoms	2.69	1.64	0.84
Number of O atoms	2.77	4.03	5.9
Number of S atoms	0.45	0.23	0.03
Number of halogen atoms	0.80	0.34	0.02
Ratio of N atoms to all heavy atoms	0.10	0.08	0.03
Ration of O atoms to all heavy atoms	0.10	0.16	0.19
Number of O + N atoms (H acceptors)	(5.5)	(5.7)	(6.8)
Number of OH + NH groups (H donors)	1.0	1.9	2.6
Number of H donors in solvated environment	**3.7**	**5.7**	**6.8**
Number of H acceptors in solvated environment	**1.0**	**1.9**	**2.6**
Number of C–C bonds	(18.5)	(15.6)	(22.5)
Number of C–N bonds	**5.3**	**3.3**	**1.9**
Number of C–O bonds	**3.1**	**4.9**	**8.2**
Number of C–S bonds	**0.75**	**0.35**	**0.05**
Number of C halogen bonds	**0.80**	**0.34**	**0.02**
SlogP	(4.3)	(2.2)	(2.4)

[a] The descriptors that show modest differences between compound classes are in parentheses.
[b] Bold descriptors are those that contributed most to the first three principal components in a PCA (principal component analysis) of the separate compound classes.

obtained by computation, namely lipophilicity (SlogP) and the number of H-bond donors and acceptors in a solvated environment. Only one descriptor is related to molecular flexibility and, hence, to the concept of property space outlined below, namely the number of rotatable bonds.

Using the 41 descriptors as independent variables in principal component analysis (PCA) allowed principal components (PCs) to be calculated, each of which is a dimension in a multidimensional space, in this case a chemical diversity space. The properties from which the first three PCs were derived are shown in bold in Table 1. The point to stress here is that each compound is but a single point in this chemical space. This should be compared with the fact that only about 66% of the variance in the datasets was explained by the first three PCs, meaning that a marked proportion of the information in the datasets was not accounted for.

Drug-likeness versus 'non-drug-likeness'

Comparable studies have attempted to assess, in a more explicit manner, the differences between drugs and non-drugs; in other words, to develop algorithmic classification techniques and, thus, aid the selection and prioritisation of compounds. Using descriptors for the whole molecule and for the presence or absence of functional groups therein, a Bayesian neural network classified as drug-like, 90% of drugs and 10% of non-drugs [12]. Another analysis based on a neural network and atom descriptors showed that drugs and non-drugs were classified adequately in 77% and 83% of cases, respectively [13]. More recent work, based on atom types and atom pair distributions, showed clear differences [14]. However, the potential problems in such analyses have recently been stressed, in particular the selection of decoy (non-drug) datasets and algorithm validation [15]. Also, it was shown that drug-like space may not necessarily be sufficiently coherent to permit a simple classification, because of non-negligible differences between therapeutic classes. In the context of this review, we note that all above analyses were based to an overwhelming extent on visually countable structural features, such as atom and moiety types, without any consideration being given to potential pharmacological relevance.

Which pharmacological criteria for drug-likeness?

For medicinal chemists involved in drug discovery, assessment of drug-likeness, based on any type of countable feature or computable property, is of little practical help. What is really needed is a profiling based on biologically relevant parameters. Here, it is important to distinguish between pharmacodynamic and pharmacokinetic events, as shown schematically in Fig. 2 [7]. Indeed,

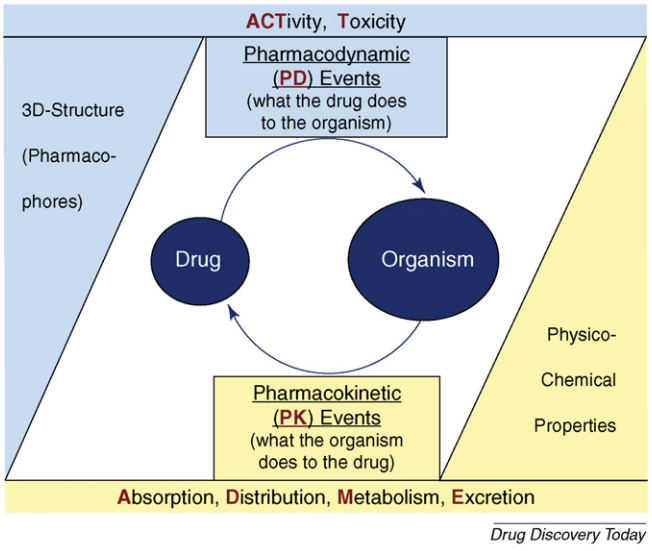

FIGURE 2

What are the relative contributions of 3D structure and phys-chem properties to PD and PK events? The figure portrays the two basic modes of interaction between xenobiotics and biological systems, namely pharmacodynamic (activity and toxicity) and pharmacokinetic events (ADME). For a bioactive compound, the biological system with which it interacts is composed of pharmacodynamic targets (receptors, ion channels, etc.) and pharmacokinetic agents (xenobiotic-metabolising enzymes, transporters, etc.). The former are often characterised by a high-ligand specificity, especially for agonists. By contrast, pharmacokinetic agents have usually evolved towards a low ligand specificity (high promiscuity), because it is their role to recognise and expel or inactivate broadly diverse substrates [7].

most pharmacodynamic events result from interaction with biological targets, such as receptors, endobiotic-metabolising enzymes, ion channels, nucleic acids, and so on. Such events are initiated when bioactive agents are recognised by (i.e. bind to) their respective target, a recognition that depends mainly, if not exclusively, on a pharmacophoric site within the protein target. In other words, pharmacodynamic events are highly dependent on 3D structure.

As for pharmacokinetic events, they involve:

- passive phenomena, such as permeation, which depend mainly on size and shape, and on lipophilicity (which encodes recognition forces such as hydrophobic interactions [hydrophobicity], H-bonding capacity and van der Waals forces [i.e. polarity]);
- recognition by proteins that have evolved to be ''promiscuous'' (i.e. to bind structurally quite diverse xenobiotics), for example drug-metabolising enzymes, xenobiotic transporters and serum proteins. Here again, molecular properties, such as hydrophobicity and H-bonding capacity, play a major role, together with some fuzzy pharmacophores defined by the presence of a small number of recognition groups and variable distances between them.

In drug discovery, the first chemical step is the physical or computational creation of ligands with high affinity for a given target, that is, fulfilling the pharmacodynamic condition. In addition, drug-like properties (the pharmacokinetic condition) are often taken into consideration as early as possible, for example, *in silico* design and combinatorial library design. Such 'drug-like' properties namely structural features and physicochemical properties that complement a given pharmacophore without affecting it, while endowing the new compounds with a potential for adequate pharmacokinetic behaviour.

Drug-likeness based on disposition

In summary, drug-likeness involves, first and foremost, molecular characteristics that are compatible with an adequate behaviour in the body, mainly good absorption, distribution and excretion. It was the great merit of Lipinski *et al.* [16] first to identify, then to list, in a formalised and simple form, the molecular properties that contribute most to drug-likeness, based on disposition criteria (Table 2).

The critical properties that emerged from the analysis are molecular weight, lipophilicity as calculated by the ClogP algorithm, and simple counts of H-bond donors and acceptors. Because each threshold is a multiple of 5, the set of criteria has been called the 'rule-of-five' (sometimes abbreviated as the Ro5). For each property, about 90% of the drugs in the database were compliant, and fewer than 10% of the drugs were non-compliant for two criteria. These cutoffs were, therefore, considered as alerts and compounds with two alerts were considered as potentially problematic; however, as made very explicit by Lipinski *et al.*, compound classes that are substrates of biological transporters are exceptions to the rule, a proviso re-emphasised more recently [17].

The Ro5 has been an inspiration for a number of informative studies which have uncovered further relevant properties (Table 2) [18–27]. Mean/median values and percentiles have been computed and differences found between leads, drug candidates and drugs. Enlightening differences were also shown to exist between drugs administered by oral, injectable or topical routes. In contrast to the Ro5, such studies do not define cutoffs. Nevertheless, their percentiles offer useful guidelines to medicinal chemists and can alert them when venturing too far from a reasonable property range. Instructive developments include recent demonstrations that target and proteomic families do influence the physicochemical profile of their ligands, in other words, that there are clear physicochemical differences between drug classes [28,29].

Based on the Ro5 and additional properties, an elaborate and promising set of 'Traffic Lights' (TLs) which addresses oral absorption has recently been proposed by Lobell *et al.* for the *in silico* prioritisation of hits [30]. Five properties have been found to be of primary importance in determining oral absorption, namely (Table 2):

- molecular size (as assessed by MW but including a correction for halogens), with optimal values $\leq$400;
- lipophilicity as calculated by the ClogP algorithm, with optimal values $\leq$3;
- solubility at pH 6.5 (i.e. resulting from the balanced contribution of neutral and ionised species), with optimal values $\geq$50 mg L^{-1};
- polarity (as assessed by the polar surface area (PSA)), with optimal values $\leq$120 $Å^2$;
- the number of rotatable bonds, with optimal values $\leq$7.

Each compound is assigned an *in silico* oral PhysChem score by summing up the values taken by its five TLs. The values of the PhysChem score can range from 0 to 10, and 'the lower the score,

TABLE 2

Rule-of-Five and some extensions as guidelines for prioritisation and optimisation

Lipinski's Rule-of-Five[a]	Additional parameters[b]	Traffic lights[c]	
MW $\leq$ 500 Da	MSA, VOL, CMR	MWcorr	$\leq$400 (TL = 0) 400–500 (TL = 1) >500 (TL = 2)
ClogP $\leq$ 5	log *D* at pH 6.5 log *D* at pH 7.4	Clog *P*	$\leq$3 (TL = 0) 3–5 (TL = 1) >5 (TL = 2)
	log Sw	Solubility (mg L^{-1})	$\geq$50 (TL = 0) 10–50 (TL = 1) <5 (TL = 2)
$\leq$5 H donors (OH and NH) $\leq$10 H acceptors (O and N)	PSA, %PSA total H-bonding count	PSA (Å^2)	$\leq$120 (TL = 0) 120–140 (TL = 1) >140 (TL = 2)
	Number of rotatable bonds	Rot bonds	$\leq$7 (TL = 0) 8–10 (TL = 1) $\geq$11 (TL = 2)
	Number of rings Number of aromatic rings Number of halogens		

[a] See text for discussion [16].
[b] Compiled from Ref. [18–27]; MSA: molecular surface area; VOL: molecular volume; CMR: calculated molecular refractivity; log S_w: calculated intrinsic water solubility; PSA: polar surface area; %PSA: percent PSA.
[c] Trafic lights (TLs) according to Lobel *et al.* [30]; MWcorr: molecular weight corrected for halogen atoms to render it proportional to molecular size; solubility: solubility at pH 6.5, calculated by an in-house expert system based on experimental data.

the more favourable the *in silico* evaluation of a compound's physicochemical properties in serving as a lead for the discovery of an orally administered drug' [30].

Property space, the missing information in current drug-likeness?

What can we conclude from the above?

As also noted for Table 1, each of the properties compiled in Table 2 takes a single value for a given compound. Together, the values of these parameters represent a compound's coordinates in a multi-dimensional property space, in a procedure that reduces each compound to a single point with inevitable loss of information. However, it is interesting to note that two types of descriptors in Table 2 are based on the implicit recognition that at least some physicochemical properties of a given compound are not single points, but cover a range of values:

- One descriptor, namely the number of rotatable bonds, stands out as containing information on a compound's conformational space. This implicit information is indirect and very limited, but it is remarkable in suggesting that conformational behaviour matters, not only in pharmacodynamic events (drug-target recognition), but also in an ADME perspective.
- Then, there are properties in Table 2 which are pH-dependent (solubility and log *D*) and will, therefore, vary for ionisable compounds depending on the nature and proportion of the electrical species present in solution. There are telling examples of failures of the Ro5 for ionised compounds [31].

In the remainder of this article, we survey the huge physicochemical and structural territory that lies beyond the door opened by these properties. This territory is the property space of a given compound. While its exploration is labour-intensive, we believe that it can be a rich source of useful information.

The concept of property space: a brief introduction and the example of acetylcholine

The concept of property space is based on the assumption that a flexible molecule (which, by definition, assumes numerous conformations) will exhibit different values for a given conformation-dependent physicochemical property. This is, in particular, the case for molecular interaction fields and, more generally, for all computable three-dimensional properties. In such cases, there will be a one-to-one relation between a given conformer and the resulting property value. The ensemble of all values a given property can take thus defines its property space. In this dynamic vision, a molecular property can be described either (a) by an average value or (b) by descriptors defining its property space (see below).

The average value of a property, and especially a weighted average, contains more information than a conformer-specific value, even if this conformer is the most probable (lowest energy) one or the bioactive one. However, this average value does not yield information on the property space itself. To this end, one should conceive and use descriptors specifying the property range and distribution over the entire conformational space, and how its variation relates to other properties.

Hence, a property space can be defined by two classes of descriptors. The first class includes descriptors which quantify the variability (spread) of values, a property range being probably the most intuitive one. The second class of descriptors relates the dynamic behaviour of a given property with other geometric or physicochemical properties. Such correlations can reveal whether and how two molecular properties vary in a coherent manner; they also lead to the concept of molecular sensitivity, because there will be sensitive molecules whose property values are markedly influenced by small conformational changes, and insensitive molecules whose properties change little even during major geometric fluctuations.

TABLE 3

The lipophilicity space of acetylcholine [32–36] in different environments, as calculated by the molecular lipophilicity potential (MLP [37–39]) for all conformers generated by long-duration molecular dynamics (MDs) simulations

Medium	Average log *P*	{Range log *P*}	{Mean RMSD} (Å)
Vacuum	−2.34	0.38	2.82
Water	−2.42	0.34	2.35
Water plus Cl^-	−2.42	0.34	2.35
$CHCl_3$	−2.36	0.35	2.37
Hydrated octanol	−2.40	0.28	1.85
Membrane model	−2.39	0.28	1.05
M_1 receptor	−2.43	0.23	0.61
M_2 receptor	−2.40	0.29	0.70
M_5 receptor	−2.39	0.27	0.66

In a series of studies [32–35], the property space of the neurotransmitter acetylcholine using molecular dynamics (MD) simulations was investigated. Acetylcholine was chosen as a model compound given its small size, flexibility and biological interest. The molecule was placed in a vacuum, in various isotropic and anisotropic media, or it was bound to human muscarinic receptors (specifically, the M_1, M_2 and M_5 subtypes) as modelled by homology techniques [36]. The results of these studies have shed light on some key features of the property space, which can be summarised as follows, taking the lipophilicity space as an example (Table 3):

- In all simulated media and within the limits of its property spaces, acetylcholine adapts its physicochemical properties to those of the environment, even if such an adaptation can occur through different mechanisms.
- Isotropic media (vacuum, water and $CHCl_3$) do not have a marked effect on the conformational space of acetylcholine, which preserves its seven conformational clusters in such media (compare the values of RMSD taken as an index of conformational freedom). Adaptability is seen in that the medium selects in each cluster the most appropriate conformers. This implies that conformational space and property spaces are not fully interdependent, and that each cluster of conformers spans most of the property space of acetylcholine. Globally, isotropic media are seen to influence only modestly the conformational and property space of acetylcholine.
- By contrast, anisotropic media, such as membrane and receptor models, vastly constrain the conformational space of acetylcholine, as seen in the total disappearance of some conformational clusters. Conversely, the lipophilicity space is not reduced proportionally, because it exhibits ranges very similar to those observed in isotropic solvents. This would suggest that anisotropic media select conformational clusters, whereas isotropic solvents select individual conformers within each cluster.
- Binding to muscarinic receptors has constraining effects, which are even stronger than those of a membrane model. Remarkably, the property space of the ligand is more constrained than its conformational space, because the receptors recognise more than one conformation of acetylcholine, while its property ranges are significantly reduced (compare with that in a vacuum). This suggests that the muscarinic receptors do not require a very narrow, highly specific bioactive conformation but rather an optimal value of some molecular property (here lipophilicity). Such optimal value is not limited to a single conformer but is characteristic of distinct conformers in various conformational clusters.
- Remarkably, range, as a descriptor of property space, appears useful to assess the mutual ligand-receptor adaptability, accounting for the (often disregarded) entropic component in receptor recognition and binding; indeed, the property ranges were able to rationalise the partial M_2 selectivity of acetylcholine [35] and to predict the α_{1a}-selectivity for a set of known adrenergic antagonists [40], suggesting that ligand selectivity is often an entropy-driven phenomenon.

Property spaces and molecular flexibility: the concept of molecular sensitivity

Although conformational and property spaces seem only partly related (see above), the results in Table 3 reveal a remarkable correlation ($r^2 = 0.83$) between the flexibility of acetylcholine (as assessed by the RMSD averages computed for each MD simulation) and its lipophilicity ranges in various media. This finding means that the ratio between molecular flexibility and property variability is seemingly constant for acetylcholine and independent of the simulated media. This also suggests that such a ratio, which encodes the ability of a molecule to modulate its properties by varying its conformational profile, is an intrinsic attribute of a given compound and can find fertile applications in quantitative descriptions of its dynamic behaviour. What is more, this ratio leads to the concept of molecular sensitivity, which has been introduced previously by considering the pair-wise correlations between physicochemical and geometric properties [40], and which we elaborate further here.

From a mathematical point of view, the molecular sensitivity for a given conformation-dependent property can be obtained as the ratio between the property range and the mean RMSD (as computed by comparing the atomic coordinates of all non-redundant conformers). In other words, molecular sensitivity can be conceived as the property range normalised by the RMSD value or, better, the property variability which a molecule can exhibit per each Å of difference in atomic positions.

On these grounds, the concept of molecular sensitivity was explored by considering a heterogeneous dataset of 125 biologically relevant molecules taken from a reference work [41]. Their conformational profiles were determined by Monte Carlo simulations. For each molecule, some physicochemical properties (namely, virtual lipophilicity by molecular lipophilicity potential (MLP) method [37,38] as implemented in the VEGA platform [39], dipole moment, PSA and solvent-accessible surface area (SAS)) were calculated for all non-redundant conformers, deriving the corresponding property space descriptors (i.e. ranges and sensitivity ratios).

The relation between the lipophilic range (values from 0 to 1.6) and molecular flexibility as expressed by RMSD averages (values from 1 to 3.5) is merely a fair one ($r^2 = 0.54$), with many compounds being far removed from the regression line (plot not shown). Compounds lying above the regression line are the sensitive

molecules, whereas compounds lying below are the less sensitive and insensitive molecules. It comes as no surprise that there is also a fair correlation ($r^2 = 0.51$) between the lipophilic range and the number of rotatable bonds (values between 1 and 16) because the number of rotatable bonds is another well-known descriptor for molecular flexibility. This suggests that the latter descriptor can also be used to derive a somewhat different definition of molecular sensitivity, namely the ratio between a property range and the number of rotatable bonds. This definition is probably more intuitive and certainly easier to compute. But whatever definition is used, the classification of molecules as sensitive or insensitive remains unchanged.

The discrimination between 'sensitive' and 'insensitive' (in fact, less sensitive) molecules becomes even more convincing when analysing a histogram of the sensitivity values as obtained from the ratio of {log *P* range} over {number of rotatable bonds}. This histogram is shown as grey columns in Fig. 3, revealing a right-skewed shape where insensitive molecules appear more abundant than the sensitive ones. Overall, the distribution ranges from 0.02 (4-ethoxyphenol, a molecule whose lipophilicity remains practically unaffected by conformational fluctuations) to 0.26 for morphine-3-*O*-glucuronide, a historical example of a 'chameleonic' molecule for its ability to adapt its lipophilicity and conformation to the medium [42].

In order to substantiate the differentiation in sensitive and insensitive molecules, the sensitivity distribution was deconvoluted in two Gaussian-type peaks using FFT filtering smoothing. The deconvolution proved successful, as confirmed by the good quality of the statistics reported in Fig. 3. As shown, it was possible to divide the dataset into two clear-cut clusters, the first of which (red line) includes the insensitive and less sensitive molecules and represents about the 60% of the dataset, whereas second peak (blue line) comprises the sensitive compounds. Specifically, the insensitive cluster is centred on a sensitivity mean equal to 0.115, while the sensitivity mean of the second cluster is equal to 0.160.

A pharmacokinetic application of molecular sensitivity

With a view to verify the utility of property space descriptors in QSAR studies and, in particular, in ADMET predictions, we extracted from the dataset a subset of 15 molecules (Table 4) for which transdermal permeability data expressed by the permeability coefficient log K_p (cm s^{-1}) were available [43]. Several studies correlate cutaneous penetration or permeation with the lipophilicity and molecular size of permeants [44], but little has been done to analyse the influence of molecular flexibility in such relationships. In a preliminary investigation, we explored the possible relations between transdermal permeability and property space descriptors for this structurally diverse subset. Interestingly, {mean log *P*} values plus molecular size (as expressed by volume) resulted in a modest correlation ($r^2 = 0.44$). By contrast, lipophilic range and sensitivity (computed from the number of rotatable bonds) together with the mean calculated log *P* produced a significantly improved correlation (Eq. (1)):

$$\log K_p = 0.58\,\{\text{mean}\log P_{MLP}\} - 2.42\,\{\text{range}\log P_{MLP}\} + 5.27\,\{\text{sensitivity}\log P_{MLP}\} - 6.87,$$
$$n = 15;\ r^2 = 0.76;\ r^2_{cv} = 0.70;\ s = 0.47;\ F = 10.93 \qquad (1)$$

Taken alone, the {mean log *P*} gave a poor correlation ($r^2 = 0.36$); adding {range log *P*} markedly improved it ($r^2 = 0.67$), and {sensitivity log *P*} further enhanced the correlation of the final relation (Eq. (1)). This result suggests that both property space descriptors (which are not cross-related, $r^2 = 0.23$) play a synergistic role in Eq. (1). Moreover, the {range log *P*} cannot be conceived as a mere

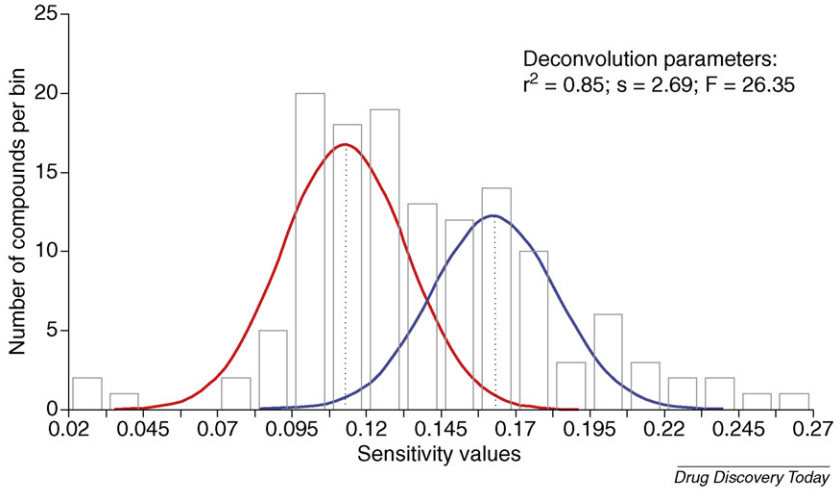

FIGURE 3

The discrimination between 'sensitive' and 'insensitive' finds a suggestive proof by analyzing a histogram of the sensitivity values as obtained from the ratio of {log *P* range} over {number of rotatable bonds}. The histogram for the dataset of 125 bioactive molecules is shown as grey columns (80 bins/unit), revealing a right-skewed shape where insensitive molecules appear more abundant than the sensitive ones. The differentiation in sensitive and insensitive molecules can be further substantiated by deconvoluting the sensitivity distribution into two Gaussian-type peaks. The deconvolution proved successful, as confirmed by the good quality of the statistics reported in figure. As shown, it was possible to divide the dataset into two clear-cut clusters, the first (red line) includes the insensitive molecules, whereas the latter (blue line) comprises the sensitive compounds.

TABLE 4

Transdermal permeability and lipophilic space parameters for a dataset of drugs [43]

Drugs	Exp log K_p	{Mean log P_{MLP}}	{Range log P}	log P sensitivity
Aspirin	−6.96	0.90	0.391	0.130
Aminopyrine	−6.62	0.85	0.102	0.051
Amylobarbitone	−6.22	2.01	0.275	0.092
Atenolol	−8.00	0.22	0.574	0.072
Atropine	−6.05	1.89	0.111	0.056
Chlorpheniramine	−6.22	3.39	0.481	0.096
Ephedrine	−5.78	1.13	0.311	0.104
Indomethacin	−5.23	3.51	0.347	0.087
Ketoprofen	−4.40	3.16	0.299	0.150
Ketorolac	−5.38	1.88	0.202	0.067
Lidocaine	−5.59	2.44	0.308	0.061
Metoprolol	−6.76	1.95	0.762	0.085
Oxprenolol	−6.37	2.51	0.758	0.084
Piroxicam	−6.11	1.98	0.341	0.114
Propranolol	−6.55	3.48	0.874	0.146

descriptor of molecular flexibility because the same relationship with {mean RMSD} *in lieu* of {range log P} affords a poorer relation ($r^2 = 0.55$).

Noticeably, the two property space descriptors appear with opposite signs in Eq. (1), suggesting that they contribute in contrasting ways to transdermal permeability. Indeed, a high range is detrimental for permeability, probably because the property variability accounts for the entropic component of permeation process. A very flexible molecule will pay too high an entropic cost because a molecule is frozen in a few suitable conformations when crossing a biomembrane. Conversely, molecular sensitivity favours transdermal permeation, probably because it encodes the adaptability that a molecule can exhibit when inserted in media of different polarity. Molecular sensitivity and adaptability thus appear as the two sides of the same coin, whose role in biological processes remains to be explored. The sensitivity ratios proposed here are the first attempt to parameterise such molecular attributes. Although the reported ADMET relationship is very far from an exhaustive analysis of the application of property space, it points to the potential of property space descriptors to account for the dynamic behaviour of bioactive compounds in quantitative structure–activity relations.

Conclusion: a plea for more molecular information

As we have argued, much molecular information is neglected in the descriptors used by medicinal chemists to assess drug-likeness. Given the confidence with which these descriptors are often used as early screening tools, one can only wonder at the proportion of potentially valuable compounds that are terminated based on incomplete or faulty information.

A missing dimension when encoding chemical information is certainly the dynamic nature of molecules. Each compound indeed spans a specific conformational space, whose range is partly controlled by the number of rotatable bonds characteristic of that compound. Following advances in computational chemistry, it is now known that molecular interaction fields (MIFs), such as molecular electrostatic potentials (MEPs), hydrophobicity and lipophilicity potentials [37,38] and H-bonding potentials [45] are conformation-dependent, as are other computable properties, such as dipole moment. A MIF such as the MLP allows a specific log P value to be computed for each conformer (also called a virtual log P), the ensemble of all such values defining the lipophilicity space of the compound. The latter is defined, for example, by the descriptor {range log P} which offers a first and partial assessment of the dynamic behaviour of a given compound. A second descriptor of potential interest is molecular sensitivity, defined as the ratio between the property range and an index of flexibility, such as the {mean RMSD}, or more simply the number of rotatable bonds.

Whatever future descriptors of dynamic behaviour will be proposed and used, they will be obtained only by further advances in our understanding of molecular properties. Our review thus ends with a plea for a greater effort in investigating the chemical properties of bioactive compounds. As shown in Fig. 4, the long road from hits (combinatorial and computational) to marketed drugs involves an enormous decrease in the number of surviving molecular entities. This attrition is compensated qualitatively by a correspondingly huge increase in the biological information generated *per* surviving compound (pharmacological and toxicological activities, pharmacokinetic and metabolic behaviour, clinical effects and interactions, to name some of the major fields of investigation). Strangely, however, there is no corresponding increase in the chemical information generated for clinical and drug candidates, in other words compounds that have passed many pharmacological, pharmacokinetic and toxicological hurdles. Much of this chemical information is obtained at the early stages of development, and it remains superficial to say the least. As we have tried to show, the more a chemical entity becomes interesting and promising, the more should be known on its molecular properties, and specifically on its dynamic behaviour and property space.

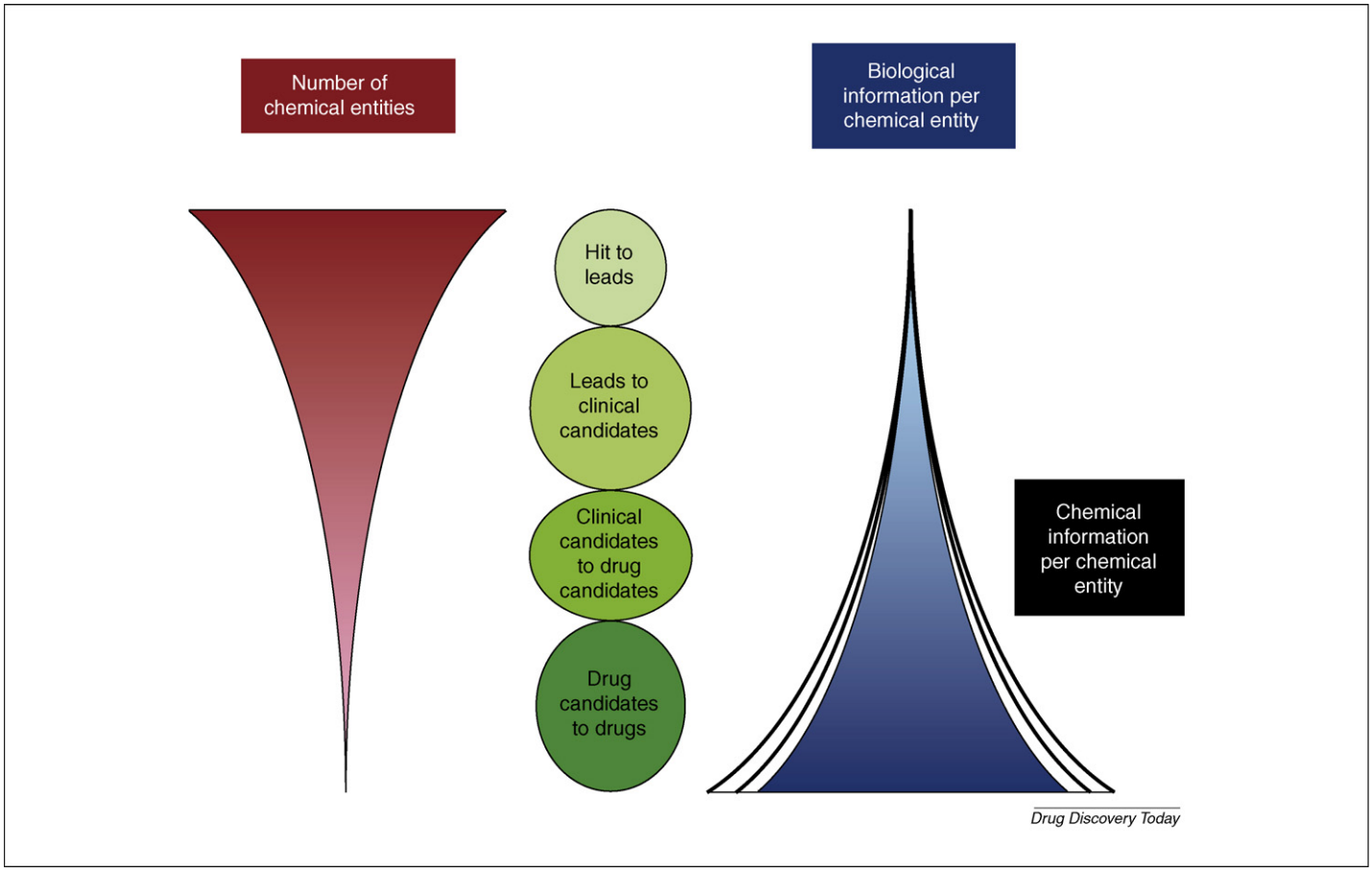

FIGURE 4

The long road from hits to marketed drugs, showing on the one hand the huge rate of attrition of chemical entities, and on the other hand the correspondingly huge increase in the biological information generated per surviving compound (in essence, clinical and drug candidates). However, the more a given chemical entity advances through development and clinical trials, the more evident its drug-like behaviour and the more valuable the information contained in its chemical structure and physicochemical properties. What the text argues for, therefore, is a corresponding increase in the chemical information obtained for clinical and drug candidates, with a view to feedback on discovery and base the drug-likeness strategy on richer and more informative descriptions of molecular properties.

References

1 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730

2 Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855–861

3 Kubinyi, H. (2003) Drug research: myths, hype and reality. *Nat. Rev. Drug Discov.* 2, 665–668

4 Cheng, A.C. *et al.* (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* 25, 71–75

5 Dawkins, R. (1989) *The Selfish Gene*. Oxford University Press

6 Leeson, P.D. and Springthorpe, B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890

7 Testa, B. *et al.* (2005) Musings on ADME predictions and structure-activity relations. *Chem. Biodivers.* 2, 1411–1427

8 Ettmayer, P. *et al.* (2004) Lessons learned from marketed and investigational prodrugs. *J. Med. Chem.* 47, 2393–2404

9 Testa, B. (2004) Prodrug research: futile or fertile? *Biochem. Pharmacol.* 68, 2097–2106

10 Testa, B. (2007) Prodrug objectives and design. In *ADME-Tox Approaches, Vol. 5 Comprehensive Medicinal Chemistry II* (Testa, B., van de Waterbeemd, H., Taylor, J.B., Triggle, D.J., eds), pp. 1009–1041, Elsevier

11 Feher, M. and Schmidt, J.M. (2003) Property distribution: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 43, 218–227

12 Ajay, A. *et al.* (1998) Can we learn to distinguish between ''drug-like'' and ''nondrug-like'' molecules? J. *Med. Chem.* 41, 3314–3324

13 Sadowski, J. and Kubinyi, H. (1998) A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329

14 Hutter, M.C. (2007) Separating drugs from nondrugs: a statistical approach using atom pair distribution. *J. Chem. Inf. Comput. Sci.* 47, 186–194

15 Good, A.C. and Hermsmeier, M.A. (2007) Measuring CAMD technique performance. 2. How ''druglike'' are drugs? Implications of random test set selection exemplified using druglikeness classification models. *J. Chem. Inf. Comput. Sci.* 47, 110–114

16 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25

17 Wu, C.Y. and Benet, L.Z. (2005) Predicting drug disposition via application of BCS: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.* 22, 11–23

18 Hann, M.M. *et al.* (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* 41, 856–864

19 Oprea, T.I. *et al.* (2001) Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* 41, 1308–1315

20 Oprea, T.I. (2002) Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput. Aided Mol. Design* 16, 325–334

21 Veber, D.F. *et al.* (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623

22 Hann, M.M. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263

23 Lajiness, M.S. *et al.* (2004) Molecular properties that influence oral drug-like behavior. *Curr. Opin. Drug Discov. Dev.* 7, 470–477

24 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232

25 Leeson, P.D. and Davis, A.M. (2004) Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem.* 47, 63386348

26 Refsgaard, H.H.F. *et al.* (2005) In silico prediction of membrane permeability from calculated molecular parameters. *J. Med. Chem.* 48, 805–811

27 Keller, T.H. *et al.* (2006) A pactical view of ''druggability''. *Curr. Opin. Chem. Biol.* 10, 357–361

28 Morphy, R. (2006) The influence of target family and functional activity on the physicochemical properties of pre-clinical compounds. *J. Med. Chem.* 49, 2969–2978

29 Vieth, M. and Sutherland, J.J. (2006) Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* 49, 3451–3453

30 Lobell, M. *et al.* (2006) In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* 1, 1229–1236

31 Martin, Y.C. (2005) A bioavailability score. *J. Med. Chem.* 48, 3164–3170

32 Vistoli, G. *et al.* (2002) The solute–solvent system: solvent constraints on the conformational dynamics of acetylcholine. *J. Am. Chem. Soc.* 124, 7472–7480

33 Vistoli, G. *et al.* (2005) Solvent constraints on the property space of acetylcholine. I. Isotropic solvents. *J. Med. Chem.* 48, 1759–1767

34 Vistoli, G. *et al.* (2005) Solvent constraints on the property space of acetylcholine. II. Ordered media. *J. Med. Chem.* 48, 6926–6935

35 Vistoli, G. *et al.* (2007) The conformational and property space of acetylcholine bound to muscarinic receptors: an entropy component accounts for the subtype selectivity of acetylcholine. *Arch. Biochem. Biophys.* 464, 112–121

36 Pedretti, A. *et al.* (2006) Muscarinic receptors: a comparative analysis of structural features and binding modes through homology modelling and molecular docking. *Chem. Biodiver.* 3, 481–501

37 Gaillard, P. *et al.* (1994) Molecular lipophilicity potential, a tool in 3D-QSAR. Method and applications. *J. Comput. Aided Mol. Design* 8, 83–96

38 Carrupt, P.A. *et al.* (1997) Computational approaches to lipophilicity: methods and applications. *Rev. Comput. Chem.* 11, 241–315

39 (2002) VEGA: a versatile program to convert, handle and visualize molecular structure on windows-based PCs. *J. Mol. Graph* 21, 47–49

40 Vistoli, G. *et al.* (2005) Range and sensitivity as descriptors of molecular property spaces in dynamic QSAR analyses. *J. Med. Chem.* 48, 4947–4952

41 Avdeef, A. (2003) *Absorption and Drug Development. Solubility, Permeability and Charge State.* Wiley–Interscience, Hoboken

42 Carrupt, P.A. *et al.* (1991) Morphine 6-glucuronide and morphine 3-glucuronide as molecular chameleons with unexpected lipophilicity. *J. Med. Chem.* 34, 1272–1275

43 Magnusson, B.M. *et al.* (2004) Molecular size as the main determinant of solute maximum flux across the skin. *J. Invest. Dermatol.* 122, 993–999

44 Geinoz, S. *et al.* (2004) Quantitative structure-permeation relationships (QSPeRs) to predict skin permeation: a critical evaluation. *Pharm. Res.* 21, 83–92

45 Rey, S. *et al.* (2001) Development of molecular hydrogen bonding potentials (MHBPs) and their application to structure-permeation relations. *J. Comput. Aided Mol. Design* 19, 521–535